Therefore, taking tangent function of both sides of (1), we get

$$\frac{\sin p \tan \alpha + \sin p \tan (\omega - \alpha)}{1 - \sin^2 p \tan \alpha \tan (\omega - \alpha)} = \tan \lambda,$$

or, by dropping functional characteristics of α, ω and λ,

$$\sin p\left(\alpha + \frac{\omega - \alpha}{1 + \alpha\omega}\right) \div \left(1 - \sin^2 p \frac{\omega\alpha - \alpha^2}{1 + \alpha\omega}\right) = \lambda. \quad . \quad . \quad . \quad . \quad (2)$$

By expansion and reduction, (2) becomes

$$(\omega \sin p - \lambda \sin^2 p)\alpha^2 - \lambda\omega \cos^2 p \alpha = \lambda - \omega \sin p. \quad . \quad . \quad . \quad (3)$$

The two roots of α in (3) are

$$\frac{\omega\lambda \cos^2 p}{2(\omega \sin p - \lambda \sin^2 p)} + \frac{[4(\lambda - \omega \sin p)(\omega \sin p - \lambda \sin^2 p) + \lambda^2\omega^2 \cos^2 p]^{\frac{1}{2}}}{2(\omega \sin p - \lambda \sin^2 p)}, \text{ and}$$

$$\frac{\omega\lambda \cos^2 p}{2(\omega \sin p - \lambda \sin^2 p)} - \frac{[4(\lambda - \omega \sin p)(\omega \sin p - \lambda \sin^2 p) + \lambda^2\omega^2 \cos^2 p]^{\frac{1}{2}}}{2(\omega \sin p - \lambda \sin^2 p)},$$

which shows, as in a plane, that the point D may be posited at the same distance from B that it is from A.

This problem, in its most general form, has a somewhat remarkable history. A general solution may be found in Newton's Universal Arithmetic, prob. 24; and in Gergonne, Annales de Mathematiques, tom. 10, p. 205.

ON ADJUSTMENT FORMULAS.

BY E. L. DE FOREST.

When a series of observed numbers shows irregularities resulting from errors of observation, the simplest mode of correcting them is by graphical construction, plotting the terms of the series on paper as ordinates to a curve, drawing a smooth curve through the points thus obtained so as to coincide with them as nearly as may seem best, and then measuring the ordinates thus corrected. This will often give results accurate enough for practical purposes. It also has the advantage of bringing the conditions of the problem together before the eye, so that graphical construction will always be a valuable adjunct and preliminary to the use of other methods which may be expected to give a higher degree of accuracy.

Of the analytical methods for adjusting a series, the simplest and most easily applied by persons of moderate mathematical knowledge is that which presupposes that the terms of the given series are equidistant, or if not, that they have been reduced to equidistant ones either graphically or by simple interpolation, and then adjusts the middle term u_0 of any group of

$2m+1$ terms by means of the formula

$$u'_0 = l_0 u_0 + l_1(u_1+u_{-1}) + l_2(u_2+u_{-2}) + \ldots + l_m(u_m+u_{-m}), \quad . \quad . \qquad (1)$$

where l_0, l_1, l_2 &c., are numerical coefficients determined in advance. It is most advantageous to assume that the true law of the series is algebraic and of a degree not higher than the third, so that if there were no errors of observation, the fourth and higher orders of differences would be zero. In point of fact, the true analytical law is usually unknown, but the above assumption respecting it will be practically safe, provided that the form of the given series, when plotted on paper is seen to approximate to that of an algebraic curve of the third degree, or a lower degree, throughout every group of $2m+1$ consecutive terms. Such a curve can have only one point of inflexion, so that the method is obviously unsuited to the adjustment of series which are known to be of sinuous form, so as to have very frequent changes of flexure.

Subject to the condition named, there may be an infinite number of systems of values assigned to the coefficients l, such that if the adjustment formula is applied to a series of the third or any lower order, the adjusted value u'_0 will be the same as the given value u_0. Schiaparelli has investigated some properties of these formulas, with a view to the reduction of meteorological observations chiefly, in his treatise *Sul modo di ricavare la vera espressione delle leggi della natura dalle curve empiriche*; Milan, 1867. He showed that the coefficients l must be connected by these two relations;

$$\left.\begin{array}{l} l_0 + 2(l_1 + l_2 + \ldots + l_m) = 1 \\ 1^2 l_1 + 2^2 l_2 + 3^2 l_3 + \ldots + m^2 l_m = 0 \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

He also ascertained what the numerical values of l must be, for values of $2m+1$ from 5 to 11 inclusive, in order that the probable errors of the adjusted terms may be reduced to a minimum. The results thus reached are really identical with those obtained by the method of least squares, when the $2m+1$ terms are each supposed to satisfy approximately the equation

$$u = a + bx + cx^2 + dx^3$$

and are also regarded as of equal weight, that is, equally liable to error. (Compare an article by the present writer, in the *Smithsonian Report* of 1871; pp. 326 and 335.) It can be shown that the adjustment formulas thus constructed are embraced under the general form

$$\begin{aligned}[(2m+1)(m^4) - 2(m^2)(m^2)]u'_0 = (m^4)u_0 &+ [(m^4) - 1^2(m^2)](u_1+u_{-1}) \\ &+ [(m^4) - 2^2(m^2)](u_2+u_{-2}) + \ldots . \\ &\ldots . + [(m^4) - m^2(m^2)](u_m+u_{-m}) \quad (3)\end{aligned}$$

where the notation used is

$$(m^n) = 1^n + 2^n + \ldots . + m^n.$$

The values of the coefficients l for all values of $2m+1$ from 5 to 25 inclusive, are here given in Table A. They are only carried out to two or three places of decimals, which are all that are needed in practice, and the last decimal figure is in some cases altered by a single unit in order that the condition $l_0 + 2(l_1 + l_2 + \dots + l_m) = 1$ may be exactly satisfied.

TABLE A.

(*giving minimum values to* $\varepsilon' \div \varepsilon$.)

$2m+1$	5	7	9	11	13	15	17	19	21	23	25
l_0	.48	.34	.26	.208	.174	.152	.134	.118	.108	.098	.090
l_1	.34	.29	.23	.196	.168	.147	.130	.117	.106	.097	.089
l_2	–.08	.14	.17	.161	.147	.133	.121	.110	.101	.093	.086
l_3		–.10	.06	.102	.112	.110	.105	.099	.093	.087	.082
l_4			–.09	.021	.063	.079	.083	.084	.081	.078	.075
l_5				–.084	.000	.038	.056	.064	.067	.067	.066
l_6					–.077	–.012	.022	.039	.049	.053	.056
l_7						–.071	—.019	.011	.027	.037	.043
l_8							—.065	—.023	.003	.019	.028
l_9								—.060	—.025	—.002	.012
l_{10}									—.056	—.026	—.006
l_{11}										—.052	—.027
l_{12}											—.049
$\frac{\varepsilon'}{\varepsilon}$	.69	.58	.50	.456	.418	.389	.365	.345	.328	.313	.300
$\frac{(\Delta'_4)}{(\Delta_4)}$	.29	.19	·13	.109	.0907	.0782	.0677	.0600	.0548	.0496	.0456

The probable error of a term adjusted by any such formula as (1), assuming that the true law of the series is algebraic and of a degree not higher than the third, will be

$$\varepsilon' = \sqrt{[l_0^2\varepsilon_0^2 + l_1^2(\varepsilon_1^2 + \varepsilon_{-1}^2) + \dots + l_m^2(\varepsilon_m^2 + \varepsilon_{-m}^2)]}$$

where ε_0, ε_1 &c., denote the probable errors of the given terms u_0, u_1 &c. If these are all equal, denoting their common value by ε we shall have

$$\frac{\varepsilon'}{\varepsilon} = \sqrt{\left[l_0^2 + 2(l_1^2 + l_2^2 + \dots + l_m^2) \right]}, \quad . \quad . \quad . \quad . \quad (4)$$

and this may be regarded as approximately the ratio which the probable error of any adjusted term will bear to that of the corresponding unadjusted one, so far as this ratio can be determined *a priori*. It is true that the errors of the given terms will not in general be really equal, but it usually

happens in statistical and other series that the weights of the terms follow a tolerably regular sequence, so that if for instance the errors ε_1, ε_2 &c., are greater than ε_0, the errors ε_{-1} and ε_{-2} will be less than ε_0, and *vice versa*, and thus they tend to correct each other. The values of l^2 too are much larger for the middle of a formula than for its extreme terms, so that the error ε' is chiefly due to the errors of u_0 and a few other terms adjacent to it. The values of $\varepsilon' \div \varepsilon$ corresponding to the several values of $2m+1$ in Table A. are given at the foot of that table. These are minimum values, for the reason that, under the conditions we have assumed, the method of least squares gives the most probable value of the adjusted term u'_0.

Before having met with Schiaparelli's work the present writer had constructed, with special reference to the adjustment of mortality tables, various systems of formulas of this general character, one of the most advantageous of which seemed to be that which renders the probable values of the fourth differences of the adjusted series a minimum. (*Smithsonian Report* of 1871, p. 332.) The series is thus graduated with the greatest possible smoothness. The coefficients under this system are here given in Table B. In a continuation of the article just cited, (Sm. Rep. of 1873, p. 327), the subject was still farther investigated. The fourth difference of any five consecutive terms in the given series being

Table B.

[*giving minimum values to* $(\Delta'_4) \div (\Delta_4.)$]

$2m+1$	5	7	9	11	13	15	17	19	21	23	25
l_0	.56	.42	.34	.290	.252	.222	.200	.182	.166	.1534	.1424
l_1	.29	.29	.27	.245	.222	.202	.184	.170	.157	.1463	.1368
l_2	-.07	.05	.11	.135	.145	.146	.143	.137	.132	.1262	.1205
l_3		-.05	-.02	.026	.056	.076	.087	.093	.096	.0969	.0962
l_4			-.03	-.028	-.007	.015	.034	.047	.057	.0635	.0678
l_5				-.023	-.027	-.018	-.004	.010	.022	.0317	.0395
l_6					-.015	-.022	-.020	-.012	-.003	.0065	.0151
l_7						-.010	-.017	-.018	-.015	-.0092	-.0024
l_8							-.007	-.013	-.016	-.0151	-.0119
l_9								-.005	-.010	-.0133	-.0140
l_{10}									-.003	-.0077	-.0110
l_{11}										-.0025	-.0060
l_{12}											-.0018
$\frac{\varepsilon'}{\varepsilon}$	.70	.60	.54	.495	.461	.433	.409	.389	.372	.358	.345
$\frac{(\Delta'_4)}{(\Delta_4)}$	.23	.080	.038	.0187	.0104	.0064	.00422	.00273	.00185	.00124	.00091

$$\Delta_4 = 6u_0 - 4(u_1 + u_{-1}) + (u_2 + u_{-2})$$

its probable value, without regard to sign, is

$$(\Delta_4) = \varepsilon\sqrt{[6^2 + 2(4^2 + 1)]} = \varepsilon\sqrt{70},$$

and hence we have

$$\varepsilon = .11952(\Delta_4). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

It was shown (p. 330) that the probable value of the fourth differences of the adjusted series is

$$(\Delta'_4) = \varepsilon\sqrt{[\Delta_0^2 + 2(\Delta_1^2 + \Delta_2^2 + \ldots. + \Delta_{m+2}^2)]}, \quad . \quad . \quad . \quad (6)$$

where Δ_0, Δ_1, Δ_2 &c., are the fourth differences of the series

$$l_2,\ l_1,\ l_0,\ l_1,\ l_2,\ l_3 \ldots. l_m,\ 0,\ 0,\ 0,\ 0.$$

Combining (5) and (6), we get

$$\frac{(\Delta'_4)}{(\Delta_4)} = .11952\sqrt{[\Delta_0^2 + 2(\Delta_1^2 + \Delta_2^2 + \ldots. + \Delta_{m+2}^2)]}, \quad . \quad . \quad . \quad (7)$$

which enables us to compute, for any adjustment formula such as (1), the ratio which the probable value of any fourth difference of the adjusted series bears to that of the corresponding fourth difference in the unadjusted series. The values of this ratio for Tables A. and B. are found at the bottom of the tables. They measure the smoothness of graduation. For instance, the ratio .00091 for the 25-term formula of Table B. shows that when that formula is used the adjusted series will be smoothly graduated to three places of decimals farther than the given series is.

An error was committed in connection with this subject, which had escaped my notice until quite recently, and I take this earliest opportunity to correct it. Since we have for the given series the relation

$$\varepsilon = .11952(\Delta_4)$$

it was too hastily inferred that we shall also have for the adjusted series

$$\varepsilon' = .11952(\Delta'_4)$$

and consequently

$$\frac{\varepsilon'}{\varepsilon} = \frac{(\Delta'_4)}{(\Delta_4)}$$

from which the conclusion was drawn that the coefficients l in Table B. (called Table III. in the article referred to) must render the ratio $\varepsilon' \div \varepsilon$ a minimum, in as much as they are known to render $(\Delta'_4) \div (\Delta_4)$ a minimum. It was thus made to appear that the formulas of Table B. are the ones which give the most probable value to the adjusted term u'_0, for any given value of $2m+1$, and that the ratio $(\Delta'_4) \div (\Delta_4)$ is, in theory at least, equal to the ratio of the probable error of the adjusted term to that of the given one. A great discrepancy between this result and the facts of the case was

indeed recognized at the time and remarked upon in the context, but it was accounted for by the existence of causes which were really of minor importance, the true cause, an error of theory, being overlooked. The truth is that, according to the theory of probabilities, the process by which we obtained the relation

$$\varepsilon = .11952(\Delta_4)$$

implies that the errors of the terms u_0, u_1, &c., of the given series are wholly independent of each other, and that errors in excess and errors in defect are equally likely to occur. This is really true for the terms of the given series, but it is not true for the adjusted series, because any five consecutive adjusted terms have been all computed, to a considerable extent, from the same given terms, only taken in somewhat different proportions, so that the errors of these adjusted terms will most probably be all in the same direction, that is, all in excess or all in defect. Hence the relation

$$\varepsilon' = .11952(\Delta'_4)$$

does not hold good, even approximately, and any deductions from it must fall to the ground. The formulas given in pages 330 and 331 of the *Smithsonian Report* of 1873 will be corrected if we substitute $(\Delta'_4)\div(\Delta_4)$ for $\varepsilon'\div\varepsilon$ where ever the latter occurs. So also the value 0.232 on page 332 really represents the ratio $(\Delta'_6)\div(\Delta_6)$, and the values 0.305 and 0.460 on page 333 represent the ratio $(\Delta'_{2+2})\div(\Delta_{2+2})$. They measure the smoothness of the adjustment, but they do not measure its accuracy. And it is not true as stated at the foot of p. 337, that when repeated adjustments are made, the ratio of the probable error of the final series to that of the original one will be the product of the ratios due to the two or more formulas used. To find the probable error in such a case, we shall have to express the final adjusted term as a linear function of the original terms, and then proceed as in the case of formulas (1) and (4).

The correct values of the ratio of error $\varepsilon'\div\varepsilon$, as found by formula (4) from the coefficients in Table B., are given at the foot of that Table. They are somewhat larger than those of Table A., while the ratios of irregularity $(\Delta'_4)\div(\Delta_4)$ are a good deal smaller. It is impossible to get at once the maximum of probable accuracy and the maximum of smoothness. Formulas may, however, be constructed so as to give intermediate and moderately small values to both ratios. Very good results are obtained by assuming that the coefficients l are ordinates to the curve

$$y = A + Cx^4 + Dx^6 + Ex^8,$$

the constant interval between these ordinates being taken as the unit of x. The values of the four constants are found from the four equations

$$\left.\begin{array}{ll}(2m+1)A+2[(m^4)C+(m^6)D+(m^8)E] & =1\\ (m^2)A+(m^6)C+(m^8)D+(m^{10})E & =0\\ A+(m+1)^4C+(m+1)^6D+(m+1)^8E & =0\\ 2C+3(m+1)^2D+4(m+1)^4E & =0\end{array}\right\} \quad . \quad . \quad . \quad . \quad (8)$$

The first two equations follow from Schiaparelli's conditions (2) already given, while the last two require that y and $dy\div dx$ should both become zero when $x=\pm(m+1)$. (Compare the *Sm. Reports* of 1871 and 1873, pp. 322 and 352.) Having obtained the numerical values of A, C, D and E for any assumed value of $2m+1$, and substituted them in the equation of the curve, we assign to x the values 0, 1, 2, ... m, in succession, and the resulting values of y will be l_0, l_1, l_2, ... l_m. These are here givea in Table C., for values of $2m+1$ from 5 to 25. The ratios $\varepsilon'\div\varepsilon$ are but little larger than those in Table A., while the ratios $(\Delta'_4)\div(\Delta_4)$ are considerably smaller, excepting for the very shortest formulas, which are of no great consequence.

TABLE C.

[*giving intermediate values to* $\varepsilon'\div\varepsilon$ and $(\Delta'_4)\div(\Delta_4)$.]

$2m+1$	5	7	9	11	13	15	17	19	21	23	25
l_0	.44	.32	.24	.206	.176	.154	.136	.122	.112	.102	.094
l_1	.37	.30	.24	.204	.176	.154	.136	.122	.112	.102	.094
l_2	-.09	.14	.18	.177	.162	.146	.133	.121	.110	.102	.094
l_3		-.10	.03	.093	.117	.121	.118	.111	.104	.098	.091
l_4			-.07	-.020	.038	.070	.085	.090	.090	.088	.085
l_5				-.057	-.037	.005	.037	.056	.067	.071	.073
l_6					-.044	-.040	—.012	.015	.034	.047	.054
l_7						-.033	—.039	—.021	.000	.018	.031
l_8							—.026	—.035	—.025	—.009	.007
l_9								—.020	—.032	—.027	—.015
l_{10}									—.016	—.028	—.026
l_{11}										—.013	—.024
l_{12}											—.011
$\frac{\varepsilon'}{\varepsilon}$	.70	.58	.50	.461	.427	.398	.375	.354	.339	.325	.311
$\frac{(\Delta'_4)}{(\Delta_4)}$	.36	.19	.088	.0516	.0324	.0207	.0146	.0103	.00802	.00589	.00480

The question as to which of the formulas of the three tables A, B or C is most advantageous for use in any given case, would be easily answered if the conditions we have assumed were really fulfilled in practice, that is to say, if the series of statistical or physical observations which we wish to

adjust were really of algebraic form and of a degree not higher than the third, subject only to accidental deviations from this law. The best formula would then be the 25-term formula of Table A., since this has the smallest ratio of error $\varepsilon' \div \varepsilon$. But actually, the true law of an observed series will probably be quite unknown, and we can only assume that a group of $2m+1$ terms will follow the algebraic law approximately, and this assumption may not be safe for so many as 25 consecutive terms. The effect of a change in the form of the function, or law of the series, may be illustrated thus:

Take the following equidistant values of the function

$$u = \sin \theta$$

which is not of algebraic form.

θ	u	θ	u	θ	u
15°	.25882	35°	.57358	60°	.86603
20	.34202	40	.64279	65	.90631
25	.42262	45	.70711	70	.93969
30	.50000	50	.76604	75	.96593
		55	.81915		

Apply to these values the 11-term formula of Table A. and the 13-term formula of Table B. The two adjusted values of the middle term sin 45° will be .70702 and .70712 which differ from the true value by .00009 and .00001.

The formula from Table B. gives the best result, although its error-ratio $\varepsilon' \div \varepsilon$ is a little the largest. The reason is mainly that l_0 and l_1 are larger, and l_m is smaller, in that formula than in the other, giving a greater weight to the middle term and terms adjacent to it. Thus the formulas of Table B. follow the actual curvature of the given series more closely than those of Table A. It seems advisable not to be governed solely by the theoretical ratios $\varepsilon' \div \varepsilon$, but to employ such formulas as appear advantageous on various grounds, and rest content if the series thus adjustsd is a smoothly graduated one, and satisfies certain tests of good adjustment, that is, certain conditions which would probably be satisfied by the true series, if we had it. There is some advantage in having a series graduated smoothly, even if the probable errors are such that two or three of the last decimal figures are in doubt, and cannot be regarded as having strictly any real value. The chief advantage is, that errors of computation can be more readily detected by the irregularity they produce in the differences of the series.

(*To be concluded in No.* 4.)